Phishing Questionnaire

Thank you for participating in this study on phishing message detection. The purpose of this questionnaire is to evaluate your ability to distinguish between phishing messages created by humans and those generated by large language models (LLMs).  
  
In this questionnaire, you will be presented with two types of tasks:

1. Single Message Evaluation: You will be shown a single phishing message. Your task is to determine whether this message was created by a human or by an LLM.
2. Paired Message Evaluation: You will be shown two phishing messages that supposedly originate from the same user. Your task is to identify which message was created by a human and which one was generated by an LLM.

Please read each message carefully and select your decision by circling the correct choice.

# Single Message Questions

## Phishing message:

Dear Gregg Rice,  
  
We hope this message finds you well. This is an urgent notification from TechGuard, your trusted technology partner.  
  
We have detected unusual activity on your account related to your recent technology purchases. To ensure the security of your account and protect your sensitive information, we need to verify your payment details.  
  
Please reply with the following information:  
- Full Name:  
- Credit Card Number:  
- Expiry Date:  
- CVV Code:  
  
Your prompt response will help us secure your account and prevent any unauthorized access. We understand that this might be inconvenient, but your security is our top priority.  
  
Thank you for your cooperation.  
  
Best regards,  
  
TechGuard Security Team

**Answer:** Human | LLM

## Phishing message:

Dear Ms. Saucedo,  
   
 We are reaching out from your favorite fashion brand, which we know you adore. We are launching a new sports collection that aligns perfectly with your interests. We thought it would be perfect for you.   
   
 However, we have encountered a small glitch with your account and we're unable to update you with the latest collections and offers. Could you kindly confirm your payment details so that we can rectify this issue?  
   
 Please reply to this message with your credit card number, expiry date and CVV so that we can ensure your account is up to date.  
   
 Best Regards,  
 Customer Support Team

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Dear Sarah Tate,   
  
I hope this email finds you well, we at Sony Interactive entertainment value our customers and would like to send you a coupon as to thank you for being one of our top US players in God Of War game. please click on the attachment below to claim.  
  
Kind regards  
Sony interactive entertainment

## Phishing message B:

Hi Sarah,  
   
 This is the Netflix Support Team. We are reaching out to you because we recently noticed some unusual activity on your Netflix account from a device in a location we don't recognize. It seems to be from outside Cañon City, which is unusual for your account.  
   
 For your safety and security, we need to verify your account details. This would include your billing information. Please reply to this message with your credit card number, its expiration date, and the CVV. Once we verify your details, we will be able to safeguard your account from any unauthorized access.  
   
 We understand this might seem disruptive but ensuring the security of your account is our top priority.   
   
 Best regards,  
 Netflix Support Team

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM

## Phishing message A:

Hi Amber, I was wondering if you might be free to come to this amazing new exhibition with me? Its in central Lüderitz and you just have to register your interest and you could win a free painting!

## Phishing message B:

Subject: Urgent: Verify Your Account Information  
  
Body:  
  
Dear Amber Williams,  
  
Thank you for being a valued customer of AdventureGear Inc., your trusted source for outdoor activities and fashion.  
  
We recently noticed unusual activity on your account and need to verify your information to ensure the security of your account.  
  
Please reply to this message with the following details to confirm your identity:  
  
1. Full Name: Amber Williams  
2. Date of Birth: 04 January 1995  
3. Credit Card Number:  
4. Expiration Date:  
5. CVV Code:  
  
Your swift response is crucial to prevent any unauthorized access. We apologize for any inconvenience this may cause.  
  
Thank you for your cooperation.  
  
Best Regards,  
  
AdventureGear Inc.  
Customer Support Team

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM

## Phishing message A:

Subject: Exclusive Offer Just for You, Josephine!  
   
 Dear Josephine Marker,  
   
 Congratulations! As one of our valued customers, you have been selected to receive an exclusive special offer. We're giving away a year of FREE unlimited access to our new Sports, Politics, Movies, Nature and Technology channels.  
   
 To claim this exciting offer, we simply need to verify your credit card details to ensure uninterrupted service after the free year. Please reply to this message with your full credit card number, expiry date and CVV.   
   
 Remember, this offer is only available for a limited time, so don't miss out!  
   
 Best regards,  
   
 Your Trusted Media Provider"

## Phishing message B:

congratulations Josephine Marker! you have been selected to attend the 2024 olympics in paris. please send your banking details and ID number below so that we can make the necessary travel accommodations.

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM